Detecting, Explaining, and Managing Anomalies in Industrial Control Systems (ICS)

Clement Fung, Eric Zeng, and Lujo Bauer
Carnegie Mellon University

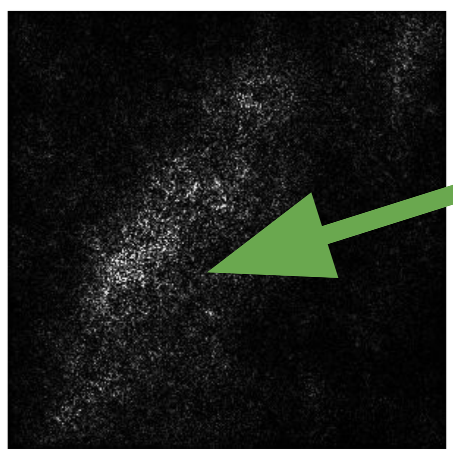
Published at
ESORICS '22
NDSS '24

Motivation

- Attacks on ICS are increasingly common
- Machine learning (ML) can be effective for anomaly detection (AD) on ICS process data
- ICS operators need **context** to diagnose alarms
- Attribution methods (e.g., saliency maps) [1] can provide context for alarms, but don't easily apply to **ICS anomaly detection**

Input Image

Attribution

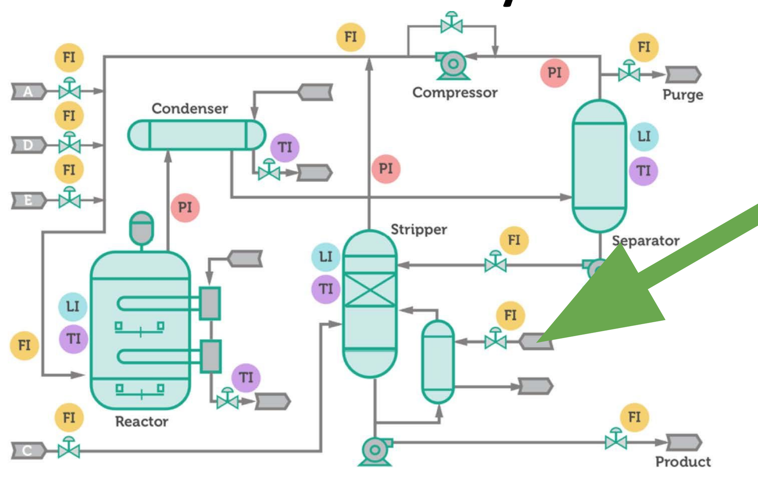


Which **input pixels** caused the prediction: "goldfinch"?

How we evaluate attribution methods

- Implement variety of ML-based AD models (CNNs, GRUs, LSTMs) [2]
- Modify chemical plant simulator to **generate 200+ attacks**
- Adapt **eight attribution methods** for AD [3]
- Analyze **properties of ICS attacks** that affect attribution accuracy [3]

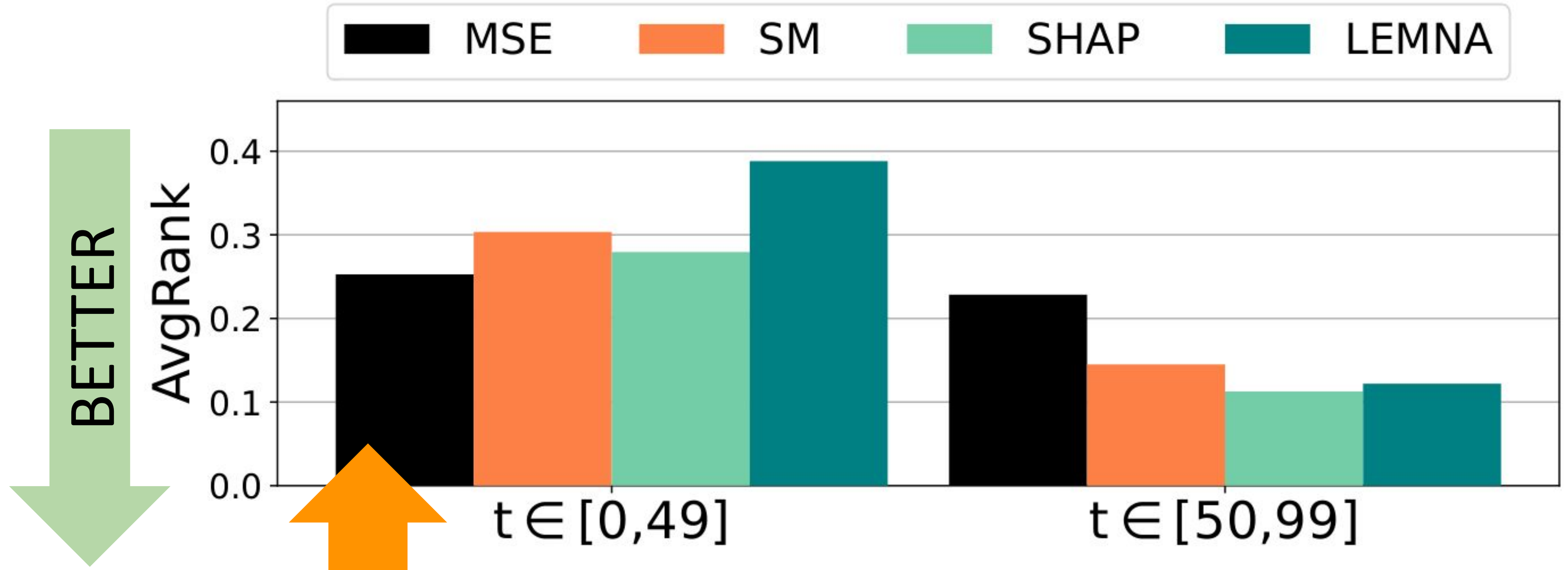
Industrial Control System



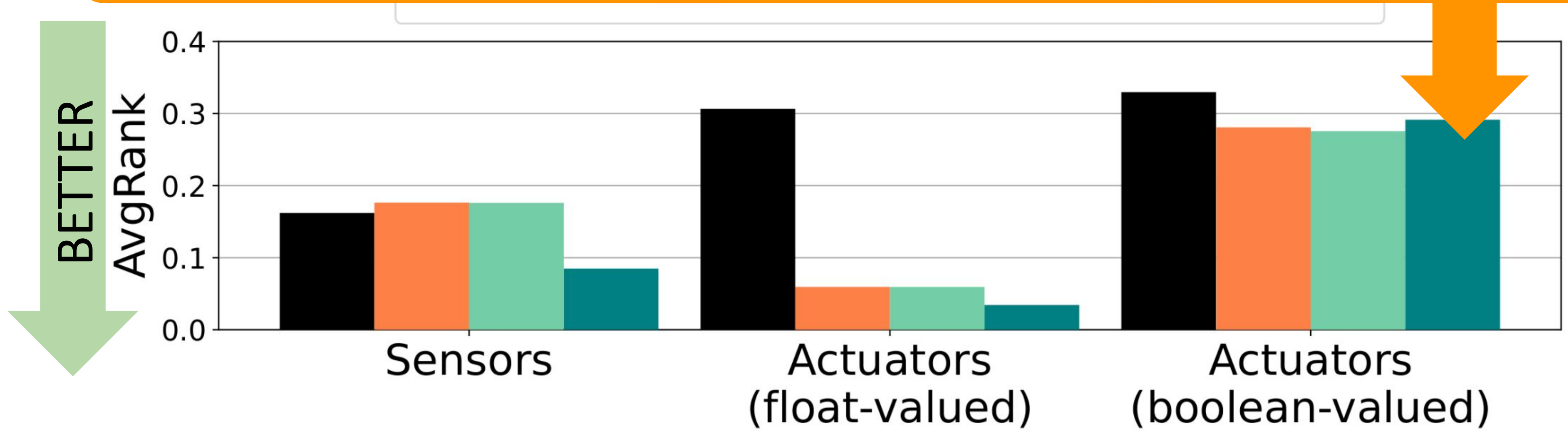
Which **sensor or actuator** was attacked?

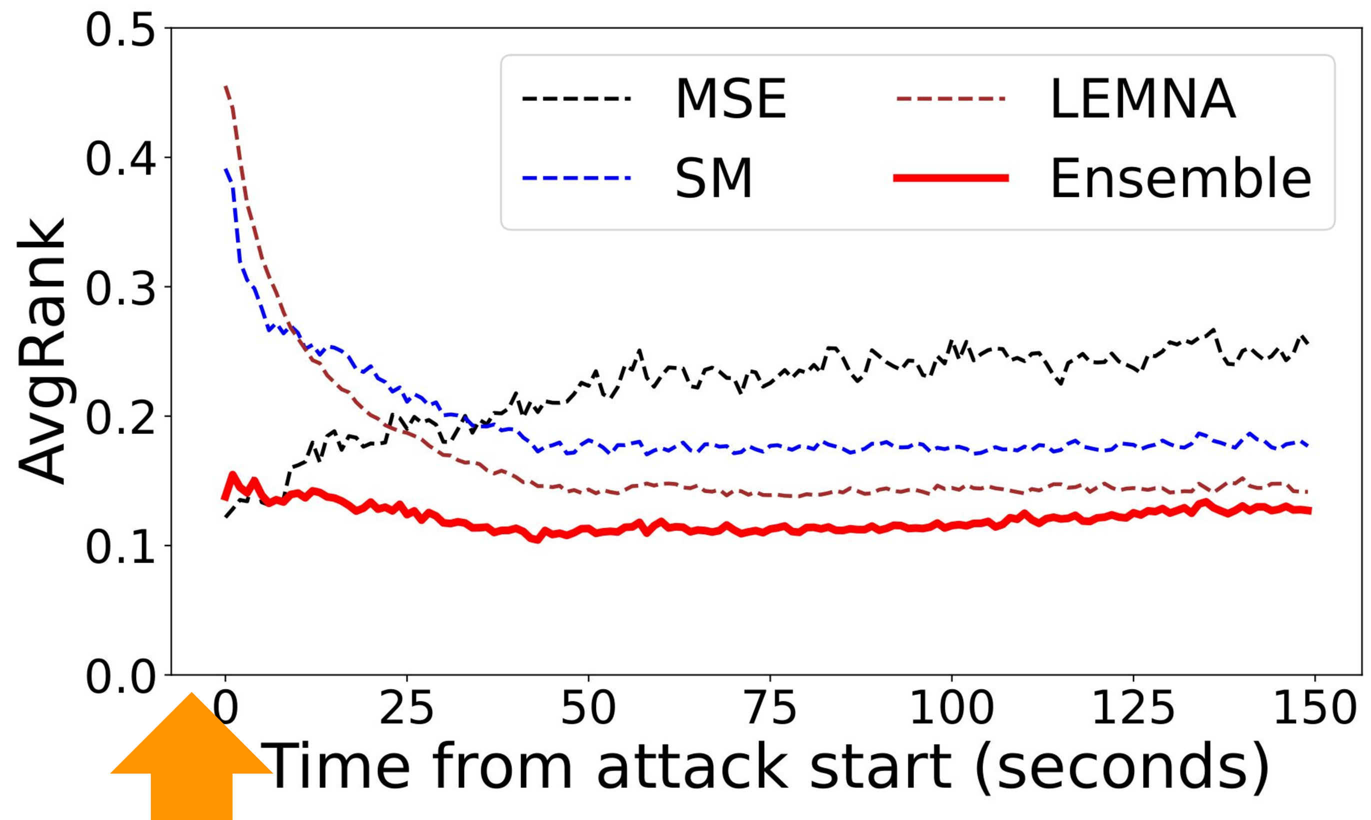
How well do attribution methods perform for ICS anomalies?

- **Finding 1**: Attribution methods (SM, SHAP, LEMNA) perform well, but depends on attack properties
- **Finding 2**: An ensemble of attribution methods outperforms any individual attribution method



Attribution is worse for: (i) early detection and (ii) boolean-valued actuators





An ensemble of attribution methods outperforms all individual methods!

What are the practical challenges for adopting ML-enabled tools for ICS?

- We interview practitioners that interact with ICS alarms: engineers, vendors, consultants, etc.
- We identify a **common alarm workflow** in ICS systems
- Challenges for adoption: technical capability, regulatory requirements, low trust in technology

Pre-incident

Set up detection system

ML for anomaly detection

Operator interface

During incident

Diagnose and respond to alarms

ML to help diagnosis (New)

Database of alarms

Post-mortem

Manage and optimize alarm system

ML for alarm optimization (New)

[1] Sturmfels et al. "Visualizing the Impact of Feature Attribution Baselines". Distill 2020.
[2] Fung et al. "Perspectives from a Comprehensive Evaluation of Reconstruction-based Anomaly Detection in Industrial Control Systems." ESORICS 2022.
[3] Fung et al. "Attributions for ML-based ICS anomaly detection: From Theory to Practice." NDSS 2024.

CyLab Carnegie Mellon University Security and Privacy Institute